Selection and Analysis of Minimal Sets of Enzyme Levels and Regulatory Structures for Optimization of Microbial Overproduction Using Large-Scale Kinetic Models of Cellular Systems

Evgeni V. Nikolaev

Department of Biomedical Engineering, Cornell University, Ithaca, NY 14853

Kew words: microbial overproduction, central metabolism, kinetic models, hybrid optimization, simulated annealing, sequential quadratic programming, control and stability analysis

July 25, 2006

Abstract

We introduce a hybrid deterministic/stochastic optimization modeling framework to identify *minimal* sets of enzyme levels and enzyme regulatory structures to meet *significant* overproduction requirements using large-scale kinetic models of microbial metabolism and essential protein machinery. Specifically, a simulated annealing algorithm is used to navigate through the discrete space of enzyme levels and regulatory structures, while a sequential quadratic programming method is utilized to identify optimal enzyme levels and regulatory kinetic parameters. The framework is demonstrated on a large-scale and chemically-detailed kinetic model of central metabolism in *Escherichia coli* (wild-type strain W3110) for the optimization of the glucose uptake through the phosphotransferase system (PTS) and serine biosynthesis. Computational results show that by optimally modulating enzyme levels and carefully altering enzyme regulatory properties, a stable 8-fold increase in the PTS uptake rate and a stable 22-fold increase in serine biosynthesis can be achieved. Importantly, substantial improvements in the targeted fluxes can be predicted by manipulating only small subsets of enzyme levels and regulatory structures. For example, the modulation of only three enzyme levels leads to a flux increase, which is almost 50% of the best predictions, and the manipulation of only six enzyme levels already leads to a flux increase of 80% of the best predictions. Remarkably, by optimally modulating 10 enzyme levels, the total central metabolism's enzyme overexpression capability is reached and any further increase in the targeted fluxes can be only possible if the pathway regulation is additionally altered, though at the expense of the loss of the pathway's steady state stability properties (i.e., no steady state can exist or oscillatory regimes may be encountered). The developed framework has also demonstrated a strong synergism between the redesign of control architectures for tightly regulated reaction steps (e.g., phosphofructokinase) and the overexpression of those enzymes which lack any type of regulatory properties (e.g., glyceraldehyde-3-phosphate dehydrogenase). Although the nonlinear optimization predictions are found in a good agreement with Metabolic Control Analysis (MCA) and large control coefficients can be indicative of the corresponding "rate limiting" enzymes and critical feedback regulatory parameters, the non-linear stable optimization predictions could not be found from the MCA alone. The proposed optimization framework thus provides a new versatile modeling strategy and computational tool for systematic optimal elucidation of minimal sets of controlling enzymes and their critical regulatory properties with broad implication in biotechnological studies.

1. Introduction

A systematic development of optimal bioprocesses and application of metabolic engineering in biotechnology and biomedical studies still requires a deeper understanding of microbial organization and function (Stephanopoulos et al. 1998; Kholodenko and Westerhoff 2004). While bioinformatics tools and related technology will continue to dominate in the filed within next decade (Overbeek et al. 2005), these efforts are, by themselves, insufficient and should be complemented by alternative tools to relate static genomes to cellular physiology and population response. One alternative way to approach this goal is to construct plausible mathematical microbial models incorporating relevant molecular and chemical details (Bailey 1998; Shuler 1999; Palsson 2004). Intrinsic complexity of cellular systems and the corresponding genomically/molecularly detailed mathematical models necessitates further development of powerful modeling concepts and computational tools to rapidly extract valuable biological information from such complex models and to provide meaningful predictive answers to practical questions arising in commercial and fundamental research.

A variety of mathematical models and modeling frameworks are already available. Similar to molecular and genetic biologists, modern modelers have now more information and tools for simulation of cellular systems with the intention of introducing desired fluxes and metabolite concentrations within complex pathways. Mathematical models have been extensively used in microbiology since the Monod's discovery of the relationship between the specific growth rate and the concentrations of limiting substrates (Monod 1949). This and similar population studies can now be complemented by credible overproduction strategies based on cellular stoichiometry alone (Burgard et al. 2003; Pharkya et al. 2003; Pharkya et al. 2004; Alper et al. 2005; Pharkya and Maranas 2006). Because stoichiometry correctly defines overall barriers and limits for intracellular and transport steady state fluxes under fixed 'defined medium' constraints, genome-scale stoichiometric models have been very successful in many instances in fundamental and applied research (Ibarra et al. 2002; Stelling et al. 2002; Almaas et al. 2004; Burgard et al. 2004). However, predictive capability of stoichiometric models is limited to calculations of 'instant phenotype snapshots' and, therefore, such models cannot capture non-stoichiometric effects mediated by isosteric and allosteric enzyme regulation (Reich and Selkov 1981), and dynamic responses in protein machinery and genetic control (Laffend and Shuler 1993; Schmid et al. 2004). To this end, alternative advances toward the rational analysis of cellular function are known as Metabolic Control Analysis (MCA) (Kacser and Burns 1973; Heinrich and Rapoport 1974) and Biochemical Systems Theory (BST) (Savageau 1969).

The MCA and *S*-systems approaches have been successfully used to improve control architectures in metabolic reaction networks (Hatzimanikatis et al. 1996; Hatzimanikatis et al. 1996). Insightful 'universal' perturbation methods have been developed to increase desired concentrations and fluxes within complex networks (Kacser and Acerenza 1993; Small and Kacser 1994). Based on this universal approach, a conception of group flux and concentration control coefficients has been introduced and then used for the optimal selection of a *small* subsets of enzymatic reactions with the *maximum* impact on the targeted flux (Stephanopoulos and Simpson 1997). An elegant Metabolic Design Analysis, based on limited kinetic and moiety conservation information, has been also suggested to answer precise 'reverse engineering' questions on the desired changes in concentrations and fluxes within complex metabolic networks (Kholodenko et al. 1998; Kholodenko et al. 2000). In this respect, a detailed analysis of genome-scale metabolic reconstructions has been performed and revealed a multitude of coupling relationships between different metabolite concentrations simultaneously present within common conserved metabolite pools (Nikolaev et al. 2005). Specifically, large conserved pools encompassing initial substrates, energy and redox cofactors, and final products can significantly hamper the applicability of the discussed perturbation approaches if such pools are ignored.

While the discussed analyses are based on linear, local linear and log-linear approximations (Heijnen 2005) of inherently nonlinear metabolic pathways, genetic manipulations typically cause metabolic networks to significantly deviate from the original stationary states or even trigger oscillatory regimes. Besides, both stoichiometric and metabolic control analyses do not provide any means to check stability of the predicted optimal states. In response to these and similar limitations and challenges, a number of research groups have undertaken the development of plausible mechanistic microbial and bacterial models. Prominent modeling projects include large-scale and chemically-detailed kinetic models of glycolysis in *Saccharomyces cerevisiae* (Rizzi et al. 1997) and central carbon metabolism in *Escherichia coli* (Chassagnole et al. 2002; Visser et al. 2004); the large-scale kinetic model of central carbon metabolism coupled with tryptophan gene expression in *E.coli* (Schmid et al. 2004); the ECell International Project (Tomita 2001); generalized coarse-grained and minimal cell models describing both metabolic and non-metabolic effects (i.e., chromosome replication, division, and cell geometry) (Shuler et al. 1979; Browning and Shuler 2001; Browning et al. 2004; Castellanos et al. 2004).

The advent of mechanistic mathematical models coupled with limitations of current modeling approaches motivates devising alternative advanced modeling concepts, computational strategies, and

frameworks to rapidly identify plausible targets for economic optimization of microbial strains. Although a lot of overproduction research has been focused on enzyme overexpression and systematic gene knockouts/insertions, radical changes in regulatory properties of metabolic networks can be often required (Stephanopoulos and Vallino 1991). While 'information flow' from DNA to RNA and then to proteins does determine hierarchically cellular function and, therefore, plays an important role in pathway activation through genetic control, however, pathway fluxes are rarely regulated by gene expression alone (ter Kuile and Westerhoff 2001), and it is really the allosteric and kinetic properties of enzymes that control the activity of those pathways (Walsh and Koshland 1985; Stephanopoulos and Vallino 1991). For example, a 50 fold overexpression of citrate synthase (EC 4.1.3.7) by recombinant DNA techniques does not increase the pathway flux (Walsh and Koshland 1985). However, changes in the concentrations of acetyl-CoA, the citrate synthase allosteric activator, and NADH, the enzyme's allosteric inhibitor, were able to considerably change this flux (Weitzman 1966; Underwood et al. 2002; Underwood et al. 2004).

Although genetic engineering techniques such as changing substrate or inhibitor affinity and recruitment of heterologous enzymes from different species is now possible (Bailey 1991; Stephanopoulos and Vallino 1991), simultaneous or consecutive enhancements of many reaction steps can lead to unpredictable instabilities or even can be experimentally infeasible (Stephanopoulos and Simpson 1997). It is therefore important to optimally select *small* numbers of most important reactions promising significant flux amplifications (Stephanopoulos and Simpson 1997). To address this and similar important problems of optimal selection and stable alteration of key enzymatic activities and kinetic regulatory properties (e.g., allosteric kinetic parameters), we have developed a general hybrid deterministic-stochastic optimization framework to systematically identify *minimal* candidate enzyme sets and their regulatory properties leading to a significant overproduction of the desired targets. Because central carbon pathways allocate the largest flux of all intracellular fluxes unevenly distributed throughout the cell (Almaas et al. 2004) and also play an important role in the providing of cellular systems with energy and *all* biosynthetic precursors (Stephanopoulos et al. 1998), a chemically-detailed kinetic model of central carbon metabolism in *E.coli* (wild-type strain W3110) (Chassagnole et al. 2002; Visser et al. 2004) has been chosen and used as a basis for benchmarking and presenting the developed optimization framework. Optimization of two important cellular functions, phosphotransferase system (PTS) and serine production, has been performed to demonstrate the applicability of the framework. Importantly, the computational modeling results have confirmed that optimal selection of "*carefully chosen small enzyme subsets*" (Stephanopoulos and

Simpson 1997) can lead to significant overproduction predictions. The calculations also revealed a strong synergism between amplification of enzyme activities of non-regulated enzymes and alteration of allosteric regulatory properties of tightly regulated committed steps in the pathway. Specifically, the simultaneous identification of two different enzyme subsets at a time is now possible, one subset, which includes enzymes subject for their subsequent overexpression in the absence of any changes in their regulation, and another subset, which includes enzymes subject for their subsequent regulatory architecture modifications without any genetically engineered changes in the specific activities. The developed framework and modeling strategy thus allow for a broad array of practically feasible genetic manipulations of a small number of cellular functions related to the targeted metabolic engineering objective.

The rest of this paper is divided as follows. In Section 2, the mechanistic kinetic model of metabolism and important constraints for essential protein machinery are formulated, and generalized regulatory structures are introduced. Section 3 presents the hybrid optimization framework and discusses its computer implementation. Section 4 and 5 present, discuss, and summarize computational optimization results, respectively.

2. The model

2.1. Modeling assumptions, constrains, and equations

A mechanistic mathematical model (1) of relevant processes in metabolism and protein machinery can be postulated as a set of kinetic mass balances coupled with synthesis/degradation equations describing the essential protein machinery and genetic control (i.e., ribosomes, RNA polymerases contents, etc) (Laffend and Shuler 1993; Schmid et al. 2004).

$$\begin{aligned} \frac{dC_i}{dt} &= \sum_{j=1}^{M} S_{ij} \cdot r_j(r_j^{\max}, \mathbf{C}, \mathbf{R}, \mathbf{K}), \quad \forall i \in \mathcal{N}, \\ \frac{de_j}{dt} &= r_j^{\text{syn}} - r_j^{\text{deg}}, \qquad\qquad \forall j \in \mathcal{M}. \end{aligned} \tag{1}$$

Here C_i is the concentration of metabolite i, e_j is the concentration of enzyme j, S_{ij} is the stoichiometric coefficient of metabolite i in reaction j with rate $r_j(r_j^{\max}, \mathbf{C}, \mathbf{R}, \mathbf{K})$ and the maximal specific reaction rate $r_j^{\max}$. $\mathbf{C}$ is the vector of metabolite concentrations, $\mathbf{R}$ is the set of all regulatory parameters (e.g., allosteric parameters), and $\mathbf{K}$ is the set of all other kinetic parameters (e.g.,

Mechaelis-Menten constants etc). Changes in enzyme levels e_j are described by the rates of enzyme synthesis r_j^{syn} and degradation r_j^{deg}. The index sets $\mathcal{N} = \{1,\ldots,N\}$, $\mathcal{M} = \{1,\ldots,M\}$, and $\mathcal{R} = \{1,\ldots,R\}$ correspond to all metabolites, reactions, and regulatory parameters, respectively.

Model (1) can be further developed for the optimal selection of an enzyme/regulatory subsystem S_D of size D such that the best possible targeted reaction rate $r_{j_0}(r_{j_0}^{\max}, \mathbf{C}, \mathbf{R}, \mathbf{K})$ can be achieved for optimized enzyme j_0. The enzyme/regulatory subsystem S_D can be constructed of the following two important subsets, E_L comprised of modulated enzyme levels and R_Q comprised of altered regulatory parameters, $S_D = E_L \bigcup R_Q$. Here $E_L = \{j_1,\ldots j_L\}$, $E_L \subseteq \mathcal{M}$, is a set of indices corresponding to the modulated enzyme levels $(e_{j_1},...,e_{j_L})$, $L \le M$, while $R_Q = \{\theta_1,\ldots, \theta_Q\}$, $R_Q \subseteq \mathcal{R}$, is a set of indices corresponding to the altered kinetic regulatory parameters $(k_{\theta_1},...,k_{\theta_Q})$, $Q \le R$. One thus has $D = L + Q$, where L and Q are the numbers of modulated enzyme levels and kinetic regulatory parameters, respectively.

Since detailed mechanistic equations describing changes in enzyme concentrations e_j are often absent, reasonable context-dependent assumptions and approximations are necessary. In this study, we follow a general approximation approach (Mauch et al. 2001), accounting for homeostasis and limited protein biosynthesis machinery in the living cell. Our goal will be to understand consequences of the introduced assumptions and approximations on the cellular physiologic response, or, more broadly speaking, to understanding how the biological mechanisms give rise to the biological function. Thinking through complex biological mechanisms and validating assumptions is a primary goal of mathematical modeling approaches, allowing for the reformulation of the original scientific problems in terms of formalized mathematical questions. Specifically, we are interested in the identification of a few important 'biotechnological' variables (i.e., enzyme levels and regulatory properties) and their causative influence on the system's overproduction properties.

First, to capture the cell's limited protein biosynthesis efforts, constraint (2) for limited changes in the ratios of enzyme masses is introduced

$$\frac{1}{M}\sum_{j=1}^{M}\frac{m_j}{m_j^0} = 1. \tag{2}$$

Here m_j and m_j^0 are the masses of enzyme j for the engineered and reference strains, respectively. Since the specific maximal reaction rate $r_j^{\max}$ is proportional to the enzyme level e_j (Stephanopoulos

et al. 1998) and $m_j = \mu_j \cdot e_j$, μ_j is the molar mass of enzyme *j*, constraint (2) can be rewritten in terms of the ratios for the maximal specific reaction rates (Mauch et al. 2001) ,

$$\frac{1}{M}\sum_{j=1}^{M}\frac{r_j^{\max}}{r_j^{\max,0}} = 1 . \tag{3}$$

Because the ratios of maximal specific reactions rates (i.e., $r_j^{\max} / r_j^{\max,0}$) are important determinants of the cellular function (Browning and Shuler 2001) and the ratios of enzyme levels (e_j / e_j^0) are readily available from the measurements (Mauch et al. 2001), constrains (2) and (3) prove very important for the modeling of metabolism coupled with essential protein machinery and genetic control.

Due to the cell's homeostasis condition (Reich and Selkov 1981; Heinrich and Schuster 1996), allowable concentration changes relative to the original stationary concentrations $\mathbf{C}_0$ should be also restricted (Mauch et al. 2001),

$$\frac{1}{N}\sum_{i=1}^{N}\frac{|C_i - C_i^0|}{C_i^0} \le \delta. \tag{4}$$

In addition to (3), constraint (4) provides further limits for the allowable reaction fluxes by avoiding unbalanced large changes in intracellular metabolite concentrations (Kacser and Acerenza 1993; Stephanopoulos and Simpson 1997). Importantly, when large changes in metabolite concentrations become incompatible with homeostasis conditions, steady states often do not exist, resulting in commonly observed undesired secretion of intermediate metabolites, impaired growth, and other unanticipated metabolic effects and events (Stephanopoulos and Simpson 1997).

Equation (1), constraints (3) and (4) are still, by themselves, insufficient to describe coordinated changes in *all* non-modulated enzymes. While molecular details on non-modulated enzyme level adjustments are currently absent, the important experimental observations show that an enormous overexpression of one enzyme in the bacterium *Zymomonas mobilis* leads to a proportion decrease in all other enzyme concentrations (Bakker et al. 1995). Based on this observation, constraints (5) are introduced to approximate the complex system response of the protein machinery network in the absence of available molecular details (Nikolaev et al. 2005) ,

$$\frac{r_{j_1'}^{\max}}{r_{j_1'}^{\max,0}} = \ldots = \frac{r_{j_K'}^{\max}}{r_{j_K'}^{\max,0}} = \gamma . \tag{5}$$

Here $j_1', \ldots, j_K'$ are the indices of non-modulated enzymes, $K = M - L$, M is the number of all enzymes

in the model (1), L is the number of modulated enzyme levels, and γ is a proportionality coefficient chosen uniform for all non-modulated enzymes. The coefficient γ can be found from (3) and (5)

$$\gamma = \frac{M - \sum_{s=1}^{L} \frac{r_{j_s}^{\max}}{r_{j_s}^{\max,0}}}{M - L}. \qquad (6)$$

Although, the control of gene expression, ribosomes, RNA polymerase, charge/uncharged tRNA contents etc is very complex, non-modulated enzyme level adjustment (5) can be interpreted as the stability of relative expression and post-translation modification of non-modulated enzymes at ratios equal to those established at the reference non-perturbed 'wild-strain' cell.

2.2. Generalized regulatory structures

Kinetic model (1) incorporates 30 enzymes and 13 enzyme regulatory structures highlighted in red in Figure 1. Because every enzyme regulatory structure is complex and encompasses many kinetic parameters (Chassagnole et al. 2002), we have implemented 13 additional dimensionless 'generalized regulatory' parameters ($\alpha_{31},...,\alpha_{43}$) to activate or disable entire regulatory structures. The model now includes 43 optimization parameters $(p_1,...,p_{43})$, $(p_1,...,p_{43}) = (r_1^{\max},...,r_{30}^{\max},\alpha_{31},...,\alpha_{43})$, where the unique numeration for all optimization parameters has been chosen (see Table 1). As a simple example of an entire regulatory structure, consider phosphoglucoisomerase (PGI) inhibition by 6-phosphoglucoconate (6pg) (Chassagnole et al. 2002)

$$r_{\mathrm{PGI}} = \frac{r_{\mathrm{PGI}}^{\max} \cdot (C_{\mathrm{g6p}} - \frac{C_{\mathrm{f6p}}}{K_{\mathrm{PGI,\,eq}}})}{K_{\mathrm{PGI,\,g6p}} \cdot (1 + \frac{C_{\mathrm{f6p}}}{K_{\mathrm{PGI,\,f6p}} \cdot (1 + \alpha_{32} \cdot \frac{C_{\mathrm{6pg}}}{1 + K_{\mathrm{PGI,\,f6p,\,6pginh}}})} + \alpha_{32} \cdot \frac{C_{\mathrm{6pg}}}{K_{\mathrm{PGI,\,g6p,\,6pginh}}}) + C_{\mathrm{g6p}}}. \qquad (7)$$

Here a dimensionless generalized regulatory parameter α_{32} is introduced to alter the regulatory properties of PGI mediated by concentration C_{6pg}. Specifically, the unit value $\alpha_{32} = 1$ corresponds to the original enzyme regulatory properties, while altered values $\alpha_{32} < 1$ (or $\alpha_{32} > 1$) correspond to decreased (or increased) regulation of PGI by $C_{6\mathrm{pg}}$, correspondingly.

3. The method

3.1. Mixed integer nonlinear problem (MINLP) formulation

To select alternative optimal targets for practically feasible enzyme level modulations and regulatory properties (e.g., inhibitory affinities) genetic mutations, optimal solutions of the following Mixed Integer NonLinear Problem (MINLP) (8) are calculated.

$$\left(\begin{array}{ll} \underset{S_D}{\text{maximize}} & r_{j_0}(r_{j_0}^{\max}, \mathbf{C}, \mathbf{R}, \mathbf{K}) \\ \text{subject to} & \sum_{j=1}^{M} S_{ij} \cdot r_j(r_j^{\max}, \mathbf{C}, \mathbf{R}, \mathbf{K}) = 0,\ i \in \mathcal{N} \\ & \frac{r_{j_1}^{\max}}{r_{j_1}^{\max,0}} + \ldots + \frac{r_{j_L}^{\max}}{r_{j_L}^{\max,0}} + K \cdot \gamma = M \\ & r_{j_s'}^{\max} = \gamma \cdot r_{j_s'}^{\max,0},\ s = 1,\ldots,K \\ & \frac{1}{N}\sum_{i=1}^{N} \frac{|\, C_i - C_i^0 \,|}{C_i^0} \le \delta \\ & \max_i \operatorname{Re} \lambda_i \le -\lambda_0 < 0 \end{array}\right) \tag{8}$$

The first constraint in (8) describes the equilibrium metabolic concentrations in kinetic mass balances (1). The second and third constraints result from the protein limited biosynthesis constraint (3) and the conditions for non-modulated enzymes (5), respectively. The forth constraint corresponds to the homeostasis condition (4). The fifth constraint explicitly encounters for the optimal solution stability properties. Here $\operatorname{Re}\lambda_i$ is the real part of eigenvalue λ_i calculated from the linearization of the right-hand side of the mass balance equation (1) at the steady state concentrations satisfying the first constraint in (8), where λ_0 is an appropriate small positive number.

Formulation (8) allows for the simultaneous elucidation of two optimal subsets E_L (i.e., a subset of modulated enzyme levels) and R_Q (i.e., a subset of altered generalized regulatory parameters) such that the best possible targeted reaction rate $r_{j_0}(r_{j_0}^{\max}, \mathbf{C}, \mathbf{R}, \mathbf{K})$ can be achieved for optimized enzyme j_0. Within the MINLP (8), the index choices E_L and R_Q are integer variables, $S_D = E_L \bigcup R_Q$, while the values of maximal specific reaction rates $(r_{j_1}^{\max},\ldots,r_{j_L}^{\max})$ and generalized regulatory parameters $(\alpha_{\theta_1},\ldots,\alpha_{\theta_Q})$ are continuous variables. Here $D = L + Q$, L is the number of modulated enzyme levels, Q is the number of altered regulatory parameters, and K is the number of non-modulated enzymes, $K = M - L$. Below a general hybrid *stochastic/deterministic* strategy to efficiently solve (8) is described in great detail.

3.2. Search for optimal subsets of enzyme levels and regulatory kinetic parameters

A simulated annealing (SA) algorithm (Kirkpatrick et al. 1983) has been implemented to navigate through the discrete space of enzyme levels $\mathcal{M}$ and generalized regulatory parameters $\mathcal{R}$ (see Figure 2). Enzyme level/regulatory subsystems S_D of increasing sizes D, $D = 1, \ldots, 10$, have been separately investigated, where $D = L + Q$, and L and Q are the numbers of modulated enzymes and altered regulatory parameters, respectively. In the pseudo-code depicted in Figure 2, S_0 is a randomly chosen initial enzyme/regulatory subsystem of size D, S_b is the enzyme/regulatory subsystem with the best reaction rate r_b found so far, S_c is the subsystem with the currently accepted reference rate r_c, and S_t is the trial subsystem with rate r_t. Parameter T is the 'annealing temperature,' reduced by factor β after each J random moves performed ($\beta < 1$), and *MaxIter* is the maximum number of all allowable SA-iterations. To generate next trial set S_t, the move class 'Select or Terminate' is implemented, where a random swap between two elements, one randomly chosen from the current subsystem S_c, $S_c \subset S_{\max}$, and another one randomly chosen from $S_{\max} \setminus S_c$, $S_{\max} = \mathcal{M} \bigcup \mathcal{R}$, is repetitively done until a new S_t is selected. The SA search is terminated when all neighbors of S_c are evaluated or the maximum number (i.e., *MaxIter*) of iterations is done.

For each set S_t, the optimal values of specific maximal reaction rates $(r_{j_1}^{\max},\ldots,r_{j_L}^{\max})$ and generalized regulatory parameters $(\alpha_{\theta_1},\ldots,\alpha_{\theta_Q})$ are calculated by utilizing standard gradient-based algorithms and techniques (e.g., an SQP-algorithm). To evaluate the given objective function $r_{j_0}(r_{j_0}^{\max}, \mathbf{C_{SS}}, \mathbf{R}, \mathbf{K})$, steady state concentrations $\mathbf{C_{ss}}$ should be known. To calculate $\mathbf{C_{ss}}$, the following two-step prediction-correction procedure was used. At the prediction step, the first kinetic mass balance equation in (1) is integrated over a fixed time span $[0, T_{\text{end}}]$. To speed up the location of steady state concentrations, the integration process is automatically terminated at an intermediate integration time moment t_{stop}, $t_{\text{stop}} \in [0, T_{\text{end}}]$, whenever the stop condition $\max_i |d\mathrm{C}_i(t_{\text{stop}})/dt| \le \varepsilon$ is met. Subsequently, at the correction step, the final integration condition $\mathbf{C}(t^*)$, $t^* = t_{\text{stop}}$ or $t^* = T_{\text{end}}$, is used as an initial guess for Newton-based solvers to find a corrected solution $\mathbf{C_{ss}}$ to the first constraint given in (8). Because Newton-based solvers converge to both stable and unstable solutions, the stability of $\mathbf{C_{ss}}$ is investigated by computing the eigenvalues of the Jacobian matrix readily available from the Newton-based solvers.

3.3. Computational implementation

The optimization framework has been demonstrated on a large-scale nonlinear model of central carbon metabolism for glucose-limited culture of *E. coli* wild-type strain W3110 (Chassagnole et al. 2002; Visser et al. 2004), comprised of 30 enzymes and 17 metabolites with the objective of maximizing fluxes through the phosphotransferase system (PTS) and serine production reaction (see Figure 1). The model includes 13 enzyme regulatory structures and the total number of all optimization parameters is 43. The values of non-perturbed specific maximal reaction rates (i.e., $r^{\max,0}$), kinetic parameters, and non-perturbed steady state metabolite concentrations (i.e., $\mathbf{C_0}$) are the same as suggested in the original publications (Chassagnole et al. 2002; Visser et al. 2004). The values of the following parameters have been fixed, $\delta = 0.1$ for allowable 10% change in metabolite concentrations (see (4)), $J = 25$, $MaxIter = 10^3$, $\beta = 0.9$, $\varepsilon = 10^{-3}$, and $T_{\text{end}} = 10^3$.

In the optimization procedure, all gradients are derived as forward finite differences, for which modulated maximal specific reaction rates are perturbed relative to the corresponding original parameter values, i.e. $r_j^{\max} = r_j^{\max,0} + \Delta r_j^{\max} = (1+h) r_j^{\max,0}$, where $\Delta r_j^{\max} = h \cdot r_j^{\max,0}$ and $h = 10^{-5}$ (Visser et al. 2004). To ensure both the robustness and the fast convergence of the SA-algorithm, different values for the initial 'simulated annealing temperature' T were used, $T_0 = 10^{-4}$ - 10^{-2} for the serine production optimization and $T_0 = 10^{-5}$ - 10^{-3} for the PTS optimal flux search. These values account for 1% - 100% of the corresponding reaction rates (mM/sec) in the reference non-perturbed steady state. The complete enumeration of all one- and two-enzymes/regulation subsystems was performed to test the ability of the algorithm to locate global optima. Also, random multistarts were used to check the robustness of the SQP search and no alternative global optima have been found (Visser et al. 2004). Since negative feedbacks stabilize the system's behavior, their attenuation during optimization search can lead to the overall destabilization of the original reaction network and, as a result, requires a tighter control of numerical errors, accompanied by additional computational cost. Specifically, tight absolute (i.e., abs = 10^{-11} - 10^{-9}) and relative (i.e., rel = 10^{-9}-10^{-7}) tolerances have been enforced to keep integration errors low due to enormous "stiffness" (i.e., parameter and initial condition sensitivity) of the original ODE model (Chassagnole et al. 2002).

We have not encountered unstable steady state solutions during the optimization search, where only enzyme levels were modulated and regulatory structures were kept unchanged (i.e., $D = L$ and $Q = 0$). In all such cases, the stability constraint in (8) was omitted to speed up calculations and, yet, the stability of optimal solutions was always checked as discussed earlier. The stability constraint was activated only if unstable solutions had been encountered. We found that unstable steady state

solutions were often encountered for mixed enzyme/regulatory subsets of large sizes D. In all such cases, the stability constraint in (8) was used with $\lambda_0 = 5 \cdot 10^{-3}$ and multistarts with different integration and optimization parameters were performed. Mathematically, the change in the stability property of stationary solutions typically correspond to the Hopf and saddle-node (fold) bifurcations (Kuznetsov 1995; Kuznetsov 2005). Specifically in our optimization studies, unstable optimal stationary solutions near the *Hopf bifurcation* were encountered for large enzyme/regulation subsets in the case of the PTS rate optimization, while unstable optimal stationary solutions near the *saddle-node bifurcation* were encountered for large enzyme/regulation subsets in the case of the serine flux optimization. Importantly, while the convergence of the root-finding Newton-based solvers near the Hopf bifurcation is always *quadratic* (i.e., the same as in general nonsingular situations), the convergence of these solvers becomes *linear* and, hence, much more slower near the saddle-node bifurcation because multiple solutions or even solutions of multiplicity two can be encountered within a small vicinity of the root (Babenko 1986).

The entire optimization modeling framework has been implemented in Matlab© on a Linux cluster with Intel CPU 3.06 GHz computers. Computational requirements were in order of minutes for small enzyme subsets S_{L+Q}, up to 30 hours for large enzyme subsets S_L (i.e., with $Q = 0$), and up to 60 hours for large enzyme/regulation subsystems S_{L+Q} (i.e., with $Q > 0$).

4. Results and discussion

4.1. Comparative analysis of optimally selected subsets

The best found flux ratios for PTS and serine (i.e., SerSynth) rates in the optimized and original models, respectively, are presented in Table 2 for the following three distinct cases, (*i*) all 13 generalized regulatory parameters are altered, while all enzyme activities are kept unchanged, (*ii*) all 30 enzyme levels are modulated, while regulatory parameters are kept fixed, and (*iii*) enzyme levels and regulatory parameters are simultaneously manipulated while the pathway stability is preserved .

We find from Table 2 that that the alteration of regulatory structures alone does not lead to significant improvements in the targeted fluxes. This can be explained by a limited capacity of an enzymatic reaction to ‘channel’ a large flux without a substantial increase in the enzyme specific activity. Indeed, if all 30 enzymes in the model are allowed to vary their levels, a 3-fold increase in the PTS rate and a 20-fold increase in the serine production can be achieved. Importantly, the optimal alteration of both specific maximal reaction rates and reaction regulatory structures leads to a 8-fold

increase in the PTS flux. At the same time, no impressive increase in the serine production flux has been observed compared to the modulation of enzyme levels alone (22.615 vs. 20.913, respectively, see Table 2). These observations are also biologically meaningful and intuitive, since PTS is a tightly regulated enzyme, while the overall lumped SerSynth 'idealized' reaction lacks any kind of pathway regulation (see Figure 1).

The analysis of small optimal enzyme/regulatory subsystems leads to the following interesting facts and conclusions. First of all, substantial improvements in the desired fluxes can be predicted by manipulating only "*small carefully chosen"* enzyme level subsets (see Figure 3) as earlier suggested by Stephanopoulos and Simpson (1997). For example, the modulation of only three enzyme levels leads to a flux increase, which is almost 50% of the best predictions, and the manipulation of only six enzyme levels already leads to a flux increase of about 80% of the best predictions (while circles in Figure 3). Importantly, by manipulating 10 enzyme levels in the absence of any regulatory changes, the central metabolism's maximum overproduction capability is reached (white circles in Figure 3, see also Table 2). We also find that no substantial increase in the desired PTS rate and serine flux can be obtained for the best *mixed* enzyme/regulation subsystems of small sizes (i.e., $D = 1$, 2, and 3) (solid circles in Figure 3), compared to enzyme level modulations alone (white circles in Figure 3).

For larger mixed enzyme/regulation choices, a significant additional flux increase can be predicted, thought at the expense of the loss of stability properties due to the attenuation of negative feedbacks, evolved to ensure robustness of cellular systems, and amplification of positive feedbacks always destabilizing the cellular system. Specifically, we could not find any stable optimal steady state solution for the serine overproduction by navigating enzyme/regulatory subsystems of size 10. Recall that, while the saddle-node bifurcation leads to the disappearance of steady state solutions, the Hopf bifurcation can give rise to stable small amplitude 'sinusoidal' periodic oscillations emanating from the original steady state which itself becomes unstable but does not disappear (Kuznetsov 1995). In the latter case, the unstable optimal stationary solution can be interpreted as averaged over the period of the small amplitude oscillations and hence, it is interesting to compare stable and unstable solutions near the Hopf bifurcation. Unstable optimal solutions can be obtained from formulation (8), where the stable constraint is previously removed. We find from Figure 3(*a*), that the optimal values for the PTS flux for both stable and unstable optimal solutions are very close (solid circles and gray diamonds in Figure 3(*a*), respectively). Despite the fact that the difference between stable and 'averaged' optimal values is small, the observation is, however, very important signifying of the emergence of potential pathway oscillations in practical biotechnological implementations

where many pathway properties (i.e., enzyme levels and regulation) are allowed to simultaneously vary. In this respect, the appearance of pathway oscillations via the Hopf bifurcation can be attributed to the well-known autocatalytic properties of glycolusis (Reich and Selkov 1981; Heinrich and Schuster 1996), enhanced by the increased glucose uptake through the optimized phosphotransferase system (PTS). In contrast, the disappearance of pathway steady states via the saddle-node bifurcation can be explained by the loss of the balance between the increased large demand for serine biosynthesis and the limited activity of the protein-making machinery which cannot support no-physiologic final product demands. All this additionally emphasizes the complexity of large-scale optimization studies and the importance of carefully chosen *small* subsets of enzyme levels and regulatory properties for their subsequent practically feasible changes and alterations (Stephanopoulos and Simpson 1997).

Remarkably, in all considered cases the calculations have demonstrated the saturation type of the optimal behavior for the entire reaction network (see Figure 3 and Table 2). This means that while more efforts can be required to elucidate larger stable optimal enzyme/regulatory choices, no further significant achievements in the corresponding targeted fluxes can be obtained independently of the stability issues and such larger choices may not be worth any significant efforts. Mathematically, the observed saturation biological behavior results from the biophysical constraints (3)-(5) introduced in the model.

4.2. Control coefficients

To get quantitative insights into how successive small enzyme/regulation subsets can be chosen to meet overproduction requirements, Metabolic Control Analysis (MCA) can be used. Originally, MCA was developed to quantify 'bottlenecks' or 'rate limiting steps' in complex pathways (Kacser and Burns 1973; Heinrich and Rapoport 1974). Taking into account that both enzyme levels and kinetic parameters (i.e., allosteric regulatory parameters) can significantly contribute to reaction rate r and overall pathway flux J, the following general definition of flux control coefficients (FCCs) has been suggested (Heinrich and Rapoport 1974; Heinrich and Schuster 1996)

$$C_r^J = \frac{r}{J} \cdot \frac{\partial J}{\partial r} = \frac{\partial \ln J}{\partial \ln r} \ . \tag{11}$$

Here J and r are a pathway flux and reaction rate, respectively, counted in their positive directions, ∂J is the infinitesimal change in J due to infinitesimal change ∂r in r under isolated conditions. Importantly, the FCCs (11) are independent of any special choice of perturbed reaction kinetic

parameters (Heinrich and Schuster 1996), given 'ideal biochemistry' conditions are met (Kholodenko and Westerhoff 1995; Heinrich and Schuster 1996). Because the model (Chassagnole et al. 2002) fulfills the ideal biochemistry conditions, the calculation of FCCs can be simplified by using infinitesimal fractional changes in specific maximal reaction rates $r^{\max}$,

$$C_r^J = \frac{r^{\max}}{J} \cdot \frac{\partial J}{\partial r^{\max}} = \frac{\partial \ln J}{\partial \ln r^{\max}}. \tag{12}$$

Although FCCs (12) provide a basis for identifying rate limiting steps, their values are, by themselves, insufficient to understand which kinetic properties (i.e., the enzyme concentration or regulation) of a rate limiting step could be important. To obtain additional relevant knowledge, the following two important coefficients can be calculated (Heinrich and Schuster 1996), (*i*) a response coefficient of flux *J* with respect to a reaction kinetic parameter α,

$$R_\alpha^J = \frac{\partial \ln J}{\partial \ln \alpha}, \tag{13}$$

and (*ii*) an elasticity coefficient quantifying the potential of parameter α to affect the reaction rate *r* under isolated conditions,

$$\pi_\alpha^r = \frac{\partial \ln r}{\partial \ln \alpha}. \tag{14}$$

The R_α^J, C_r^J, and π_α^r satisfy the following simple identity (Heinrich and Schuster 1996; Fell 1997) when the ideal biochemistry conditions are met, where $\pi_\alpha^r \equiv 1$ and, hence, $R_\alpha^J \equiv C_r^J$ for $\alpha = r^{\max}$,

$$R_\alpha^J \equiv C_r^J \cdot \pi_\alpha^r. \tag{15}$$

The identity (15) can be interpreted as 'symbiosis' of specific features of isolated enzymes and 'networking' properties of complex pathways. Using a log-log finite-difference approximation of the left and right hand sides in (15), the identity (15) has been numerically checked and the corresponding coefficients have been obtained (see Table 3 and Figure 4). We find from Table 3 that, for example, the two coefficients R_α^J calculated for PEP carboxylase (PEPxylase) signify of the opposite effects on the pathway optimal response caused by the same relative infinitesimal perturbation of enzyme regulatory properties, R_α^J is positive for the PTS flux and is negative for the serine flux. For our practical optimization studies, this means that the role of specific enzyme regulatory properties depends on the network targeted flux selected for the practical optimization.

Importantly, the calculated control coefficients reveal a distributed control on the targeted fluxes, allocated within a group of several rate limiting steps exerting the highest control. Namely, the same group of rate limiting steps (i.e., PTS, PFK, GAPDH, PK, PDH, PEPxylase, and G6PDH) is identified for potential practical enzyme level modulations and regulatory structures genetic mutations for both cases of the PTS and serine optimizations. Since the control coefficients are readily available from the measurements (Fell 1997) , we will compare local MCA-based predictions with those obtained from the proposed optimization framework. Specifically, the following frequently asked important questions will be addressed and guide our comparative analysis.

- Which minimal sets of enzymatic reactions and their regulatory properties should be selected to get the maximum stable impact on the targeted flux?
- Whether only enzymes exerting the highest control should be perturbed or near-equilibrium enzymes with negligibly small control coefficients are also important?
- Whether both level and regulation of the same enzyme (e.g., a committed enzyme in the pathway) should be simultaneously manipulated or different groups of enzymes could be selected so that the enzyme activity and regulatory properties could be separately modified within each group of enzymes?

4.3. PTS rate optimization

To facilitate the analysis of the best enzyme/regulation choices leading to a substantial increase of the glucose uptake through the phosphotransferase system (i.e., the PTS rate) and to compare these choices with the MCA-based predictions, the best enzyme/regulation sets are organized within Table 4. The practical importance of both enzyme level overexpression and enzyme regulatory properties genetic mutations necessitates a comparative discussion of theses two cases obtainede from separate calculations.

We begin with the presenting and discussion of the best enzyme level choices. First of all, the best enzyme level choices are found to be in a good agreement with the MCA-based predictions, though there are some non-intuitive discrepancies. Specifically, the PTS level (i.e. **1(+)** in Table 4) is always suggested to be modulated and this is in a complete agreement with biological intuition. However, the detailed analysis of the results presented in Table 4 reveals that both local FCC-based and nonlinear optimization predictions lack the *additivity* property in a sense that the best enzyme choices alone cannot be combined one another to significantly improve the PTS rate. For example,

the triplet of enzymes PTS (**1**), phosphofructokinase (PFK) (**3**), and pyruvate dehydrogenase (PDH) (**11**) exerting the highest total control (see Figure 4) is absent from Table 4. These enzymes are, however, present in all larger subsets (i.e., with $D > 4$). Importantly, while MCA suggests decreasing the level of glyceraldehyde-3-phosphate dehydrogenase (GAPDH) exerting high negative control (see Figure 4), the level of this enzyme is increased in all nonlinear optimization studies (i.e. **6(+)** in Table 4). This observation is also confirmed and is discussed in a similar optimization study by Visser at al. (2004). Therefore, even for a rate limiting step with a high control coefficient, the direction of the corresponding enzyme level modulation cannot be solely predicted based by the MCA alone.

The optimization results also show how the best enzyme level choices emerge. Although, the best choices lack the additivity property, the best smaller choices repeatedly enter the best larger subsets (see Table 4). This means that control of flux does not shift between different groups of enzymes due to the compensating effects of global regulation and homeostasis. The absence of shift in distributed control additionally emphasizes the importance of rate limiting steps with high values of control coefficients (Stephanopoulos and Simpson 1997).

Enzymes with large flux control coefficients are not always the ones to be modified, especially if they are involved in feedback control loops (Kacser and Burns 1973; Heinrich and Rapoport 1974). It may be the removal or attenuation of certain negative-feedback loops that should be considered and not the amplification of the activity of the corresponding enzymes (Stephanopoulos and Vallino 1991). These and similar important theoretic predictions are exemplified by the optimal selection of the best *mixed* enzyme level/regulatory choices, presented in the right side of Table 4. Importantly, the regulatory properties and not enzyme levels of all three tightly regulated enzymes (i.e., PTS **(1:31)**, PFK (**3:33-34**), and PEPxylase (**12:43**)) have been chosen to alter, while their levels were automatically adjusted due to the limited protein machinery constraints (3) and (5). These nonlinear optimization observations are found in a good agreement with calculations of response coefficients presented in Table 3. For example, the attenuation of the negative regulation of PFK by pep (**3:33**) (FRC = -0.496) is more preferable than the amplification of the positive feedbacks by amp (**3:35**) (FRC = 0.065), and adp (**3:34**) (FRC = 0.00605), see Table 3. Remarkably, the same enzymes are present in the left and right sides of Table 4, (see the first indices in the double notation ($j : s$)), while different manipulations of these enzymes are suggested, based on the presence or absence of regulatory structures in the selected enzymes (see Table 4).

Figure 5 compares flux control coefficients and relative optimal enzyme levels, calculated for the six best enzymes optimally selected to increase the PTS flux (i.e., PTS (1), PFK (3), TIS (5), GAPDH (6), PDH (11), and PEPxylase (12)). The corresponding distributions of steady state fluxes are shown in Figure 6. Importantly, the levels of modulated enzymes (see Figure 5(*b*)) are proportionally changed accordingly to the changes in their flux control coefficients (see Figure 5(*a*)). Specifically, the change in the sign of the FCCs (i.e., negative values are changed to positive values) for GAPDH correctly predicts a substantially increased level of the enzyme when both the enzyme levels and regulation of all selected enzymes are allowed to vary. Interestingly, a near-equilibrium enzyme TIS requires about a 3-fold increase (and not decrease) in its level despite the fact that its FCC is negligibly small. Importantly, the simultaneous manipulations of both enzyme levels and regulations led to a more 'even' redistribution of the total enzyme protein between modulated and non-modulated enzymes. Specifically, when all six enzyme levels are only modulated, a substantial uniform decrease by factor 0.27 (i.e., $\gamma = 0.27$ as defined in (5) and (6)) in the levels of the non-modulated enzymes was required, while a more moderate 0.59-fold decrease in the levels of the non-modulated enzymes (i.e., $\gamma = 0.59$) was enough when optimal changes in both enzyme levels and regulations were introduced. We also found that the inhibition of PTS by glucose-6-phosphate (g6p) was decreased by factor 0.054 (i.e., $\alpha_{31} = 0.054$), the inhibition of PFK by phosphoenolpyruvate (pep) was attenuated by factor 0.07 (i.e., $\alpha_{33} = 0.07$), and the activation of PEP carboxylase (PEPxylase) by fructose-1,6-bisphosphate (fdp) was increased by factor 161.3 (i.e., $\alpha_{43} = 161.3$).

4.4. Serine production optimization

Similar results have been obtained for the optimization of the serine production flux, see Table 5. We note that small enzyme/regulation choices are, as well, intuitive as the PTS (**1+**) supplies metabolism with the initial substrate, while SerSynth (**17(+)**) leads to the final serine production. Large enzyme/regulation choices encompass also enzymes with both high and negligibly low values of FCCs. Remarkably, the level of PDH (i.e. **11(+/-)** in the right side of Table 5) with a relatively high *positive* control on the serine flux (see Figure 4) is increased for smaller enzyme sets and is decreased for larger enzyme sets when the pathway regulation is allowed to vary. This observation could not be predicted from the analysis of the best FCCs alone and can be presumably explained by the saturation of the PDH activity, when no more overexpressed quantities of this enzyme are required.

Because of the importance of the robustness and stability issues for practical applications (Stephanopoulos and Simpson 1997), we have chosen an intermediate stable case, comprised of the best six enzymes to demonstrate optimal enzyme activities and regulatory properties. Recall that this case already provides an increase in the serine production of about 80% of the best predictions shown in Figure 3(*b*) (white and solid circles). We find that an increase in the serine demand (see Figure 7(*b*)) reallocates the strength of metabolic control from the serine synthesis (SerSynth in Figure 7(*a*)) towards the supply block (i.e. PTS) and the pyruvate removal block (i.e. PDH) (see Figure 7(*a*)). The optimal distributions of the steady state fluxes allocated within the pathway towards the serine overproductions are shown in Figure 8. These nonlinear optimization observations are also found in a good agreement with calculations of response coefficients (RFC) presented in Table 3. However, the amplification of the positive regulation of PFK by amp (FRC = 0.0376) (**3:35**) has proved to be more preferable than the attenuation of negative feedback by pep (FRC = -0.287) (**3:33**).

In this case, the inhibition of PTS by g6p was decreased by factor 0.057 (i.e., $\alpha_{31} = 0.057$), which is similar to the factor obtained in the case of the PTS optimization, and the activation of PFK by amp was increased by factor 11.74 (i.e., $\alpha_{35} = 0.07$). We also find that similar optimal adjustment γ-factors for the levels of non-modulated enzymes have been calculated from (5) for both cases, where regulatory structures are kept unchanged ($\gamma = 0.16$) or these are allowed to vary ($\gamma = 0.19$). This is because in both cases the high optimal levels of the lumped 'SerSynth' enzyme are required to meet significant overproduction objective, see Figure 7(*b*), while the levels of the other enzymes should be adjusted accordingly to the cell's limited protein biosynthesis efforts, see (3) and (5).

Thus, the computational optimization results confirm the importance of high flux control coefficients estimated at the reference non-perturbed ('wild-type') strain, which correctly delineate the most important blocks of central metabolism from less important subordinate pathways (i.e. other biosynthetic routes). Specifically, flux is increased through phosphotransferase system (PTS), phosphofructokinase (PFK), a committed enzyme in the network, and pyruvate dehydrogenase (PDH) to remove an excess of pyruvate accumulated through the enhanced PTS.

Comparing the best enzyme choices for the optimal PTS rates and serine fluxes, we find that in both cases the best choices emerge in a similar fashion signifying of the common trend in the selection of candidate enzyme/regulation subsets, where the best smaller choices repeatedly enter the best larger subsets (see Tables 4 and 5). This remarkable observation can be used to customize the simulating annealing-based random search, where small best enzyme choices can be used as initial

sets for the optimal selection of larger subsets and the random search can be biased toward more frequent selection of enzyme levels and regulatory parameters exerting the highest control.

5. Conclusions

A general hybrid stochastic/deterministic optimization framework for optimal selection of enzyme levels and regulatory structures using large-scale kinetic models of cellular systems has been introduced and demonstrated on the dynamic large-scale model of central carbon metabolism of *Escherichia coli* (Chassagnole et al. 2002; Visser et al. 2004). A simulated annealing algorithm was employed to navigate through the discrete space of enzyme levels and enzyme regulatory structures, while general gradient-based search methods were used to estimate optimal values of enzyme levels and regulatory kinetic parameters. The proposed framework allows for the optimization of the entire cellular system where by systematically selecting *small* enzyme/regulation sets and subsystems, feasible for experimental implementations, significant many-fold production improvements can be predicted. Although, the developed optimization approach does not guarantee that the optimal solutions obtained are global and better choices can be still found, the optimal solutions discussed already provide valuable candidate enzyme/regulation choices, which may be useful in prioritizing theoretic and practical studies of important properties of enzymatic reactions, kinetic regulatory structures, and providing a systematic framework for designing experiments to better understand regulation of cellular function. Remarkably, the framework can be also used as a powerful computational theoretic tool for the direct validation of modeling predictions and context-dependent theoretic assumptions allowing the modeler to understand how the biological mechanisms give rise to the biological function.

Alternatively, the optimization framework can be used in biomedical studies to identify enzymes controlling undesired large metabolite concentrations and fluxes, and to establish metabolic dysfunction using mathematical models, which otherwise would be very difficult to detect from experimental and microarray studies alone (Schuster and Holzhutter 1995; Stephanopoulos and Simpson 1997). Specifically, given an experimentally detected abnormal increase in an important decease biomarker (Bandara et al. 2003), the optimization framework can be used as a ‘reverse-engineering’ approach to generate alternative sets of candidate enzymes and regulatory structures, spontaneous toxic- or aging-related mutations in which, could be responsible for causing the symptoms of human disease. Biomedical expertise then could be employed to narrow down the

search, based on the computationally selected 'optimal' enzyme level/regulation choices undesired under such decease-related circumstances.

Nomenclature

Enzymes

ALDO	aldolase
DAHPS	DAHP synthases
ENO	enolase
G1PAT	glucose-1-phosphate adenyltransferase
G3PDH	glycerol-3-phosphate dehydrogenase
G6PDH	glucose-6-phosphate dehydrogenase
GAPDH	glyceraldehyde-3-phosphate dehydrogenase
MetSynth	methionine synthesis
MurSynth	mureine synthesis
PFK	phosphofructokinase
PGDH	6-phosphogluconate dehydrogenase
PGI	glucose-6-phosphate isomerase
PGK	phosphoglycerate kinase
PGM	phosphoglycerate mutase
PDH	pyruvate dehydrogenase
PEPxylase	PEP carboxylase
PGlucoM	phosphoglucomutase
PK	pyruvate kinase
PTS	phosphotransferase system
R5PI	ribose-phosphate isomerase
RPPK	ribose-phosphate pyrophosphokinase
Ru5P	ribulose-phosphate epimerase
Synth1	synthesis1
Synth2	synthesis2
TA	transaldolase
TIS	triosephosphate isomerase

TKa	transketolase A
TKb	transketolase B
TrpSynth	tryptophan synthesis

Metabolites

2pg	2-phosphoglycerate
3pg	3-phosphoglycerate
6pg	6-phosphogluconate
accoa	acetyl-coenzyme A
dhap	dihydroxyacetonephosphate
e4p	erythrose-4-phosphate
f6p	fructose-6-phosphate
fdp	fructose-1,6-bisphosphate
g1p	glucose-1-phosphate
g6p	glucose-6-phosphate
gap	glyceraldehyde-3-phosphate
glc	glucose
oaa	oxaloacetate
pep	phosphoenolpyruvate
pgp	1,3-diphosphoglycerate
pyr	pyruvate
rib5P	ribose-5-phosphate
ribu5p	ribulose-5-phosphate
sed7p	sedoheptulose-7-phosphate
xyl5p	xylulose-5-phosphate

References

Almaas, E., B. Kovacs, T. Vicsek, Z. N. Oltvai and A. L. Barabasi (2004). "Global organization of metabolic fluxes in the bacterium *Escherichia coli*." Nature **427**(6977): 839-43.

Alper, H., Y. S. Jin, J. F. Moxley and G. Stephanopoulos (2005). "Identifying gene targets for the metabolic engineering of lycopene biosynthesis in Escherichia coli." Metab Eng **7**(3): 155-64.

Babenko, K. I. (1986). The Foundation of the Numetical Analysis. Moscow, Nauka (*in Russian*).

Bailey, J. E. (1991). "Toward a science of metabolic engineering." Science **252**(5013): 1668-75.

Bailey, J. E. (1998). "Mathematical modeling and analysis in biochemical engineering: past accomplishments and future opportunities." Biotechnol Prog **14**(1): 8-20.

Bakker, B. M., H. V. Westerhoff and P. A. Michels (1995). "Regulation and control of compartmentalized glycolysis in bloodstream form Trypanosoma brucei." J Bioenerg Biomembr **27**(5): 513-25.

Bandara, L. R., M. D. Kelly, E. A. Lock and S. Kennedy (2003). "A potential biomarker of kidney damage identified by proteomics: preliminary findings." Biomarkers **8**(3-4): 272-86.

Browning, S. T., M. Castellanos and M. L. Shuler (2004). "Robust control of initiation of prokaryotic chromosome replication: essential considerations for a minimal cell." Biotechnol Bioeng **88**(5): 575-84.

Browning, S. T. and M. L. Shuler (2001). "Towards the development of a minimal cell model by generalization of a model of Escherichia coli: use of dimensionless rate parameters." Biotechnol Bioeng **76**(3): 187-92.

Burgard, A. P., E. V. Nikolaev, C. H. Schilling and C. D. Maranas (2004). "Flux Coupling Analysis of Genome-scale Metabolic Reconstructions." Genome Research **14**: 301-312.

Burgard, A. P., P. Pharkya and C. D. Maranas (2003). "OptKnock: A Bilevel Programming Framework for Identifying Gene Knockout Strategies for Microbial Strain Optimization." Biotechnology and Bioengineering **84**: 647-657.

Castellanos, M., D. B. Wilson and M. L. Shuler (2004). "A modular minimal cell model: purine and pyrimidine transport and metabolism." Proc Natl Acad Sci U S A **101**(17): 6681-6.

Chassagnole, C., N. Noisommit-Rizzi, J. W. Schmid, K. Mauch and M. Reuss (2002). "Dynamic modeling of the central carbon metabolism of Escherichia coli." Biotechnology and Bioengineering **79**(1): 53-73.

Fell, D. (1997). Understanding the Control of Metabolism. London and Miami, Portland Press.

Hatzimanikatis, V., C. A. Floudas and J. Bailey (1996). "Analysis and Design of Metabolic Reaction Networks via Mixed-Integer Linear Optimization." AIChE Journal **42**(5): 1277-1292.

Hatzimanikatis, V., C. A. Floudas and J. Bailey (1996). "Optimization of regulatory Architectures in metabolic reaction networks." Biotechnology and Bioengineering **52**: 485-500.

Heijnen, J. J. (2005). "Approximative kinetic formats used in metabolic network modeling." Biotechnol Bioeng **91**(5): 534-45.

Heinrich, R. and T. A. Rapoport (1974). "A linear steady-state treatment of enzymatic chains. General properties, control and effector strength." Eur J Biochem **42**(1): 89-95.

Heinrich, R. and S. Schuster (1996). The Regulation of Cellular Systems. New York, Chapman & Hall.

Ibarra, R. U., J. S. Edwards and B. Ø. Palsson (2002). "Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth." Nature **420**: 186-189.

Kacser, H. and L. Acerenza (1993). "A universal method for achieving increases in metabolite production." Eur J Biochem **216**(2): 361-7.

Kacser, H. and J. A. Burns (1973). "The Control of flux." Symp Soc Exp Biol **27**: 65-104.

Kholodenko, B. N., M. Cascante, J. B. Hoek, H. V. Westerhoff and J. Schwaber (1998). "Metabolic design: how to engineer a living cell to desired metabolite concentrations and fluxes." Biotechnol Bioeng **59**(2): 239-47.

Kholodenko, B. N. and H. V. Westerhoff (1995). "The macroworld versus the microworld of biochemical regulation and control." Trends Biochem Sci **20**(2): 52-4.

Kholodenko, B. N. and H. V. Westerhoff (2004). Metabolic engineering in the post-genomics era. Wymondham, UK, Horizon Bioscience.

Kholodenko, B. N., H. V. Westerhoff, J. Schwaber and M. Cascante (2000). "Engineering a living cell to desired metabolite concentrations and fluxes: pathways with multifunctional enzymes." Metab Eng **2**(1): 1-13.

Kirkpatrick, S., C. D. J. Gelatt and M. P. Vecchi (1983). "Optimization by simulated annealing." Science **220**(4598): 671-680.

Kuznetsov, Y. A. (1995). Elements of Applied Bifurcation Theory. New York, Springer.

Kuznetsov, Y. A. (2005). Trends in Bifurcation Software: From CONTENT to MATCONT. 4th Workshop on Computation of Biochemical Pathways and Genetic Networks. A BioSim Event. U. Kummer, J. Pahle, I. Surovtsova and J. Zobeley. Villa Bosch, Heidelberg**:** 49-57.

Laffend, L. and M. L. Shuler (1993). "Ribosomal Protein Limitations in *Escherichia coli* under Conditions of High Translational Activity." Biotechnology and Bioengineering **43**: 388-398.

Mauch, K., S. Buziol, J. Schmid and M. Reuss (2001). Computer-Aided Design of Metabolic Networks. Chemical Process Control-6 Conference, Tucson, Arizona.

Monod, J. (1949). "The growth of bacerial cultures." Ann. Rev. of Microbiol. **111**: 371-394.

Nikolaev, E. V., A. P. Burgard and C. D. Maranas (2005). "Elucidation and Structural Analysis of Conserved Pools for Genome-Scale Metabolic Reconstructions." Biophysical Journal **88**(1): 1-13.

Nikolaev, E. V., P. Pharkya, C. D. Maranas and A. Armaou (2005). Optimal Selection of Enzyme Levels Using Large-Scale Kinetic Models. 16-th International Federation of Automatic Control (IFAC) World Congress in Prague, Prague, Czech Republic (*to appear*).

Overbeek, R., T. Begley, R. M. Butler, et al. (2005). "The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes." Nucleic Acids Res **33**(17): 5691-702.

Palsson, B. O. (2004). "In silico biotechnology. Era of reconstruction and interrogation." Curr Opin Biotechnol **15**(1): 50-1.

Pharkya, P., A. P. Burgard and C. D. Maranas (2003). "Exploring the overproduction of amino acids using the bilevel optimization framework OptKnock." Biotechnology and Bioengineering **84**: 887-899.

Pharkya, P., A. P. Burgard and C. D. Maranas (2004). "OptStrain: a computational framework for redesign of microbial production systems." Genome Res **14**(11): 2367-76.

Pharkya, P. and C. D. Maranas (2006). "An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial systems." Metab Eng **8**(1): 1-13.

Reich, J. G. and E. E. Selkov (1981). Energy Metabolism of the Cell. A Theoretical Treatise. New York, Academic Press.

Rizzi, M., M. Baltes, U. Theobald and M. Reuss (1997). "In vivo analysis of metabolic dynamics in Saccharomyces cerevisiae: II. Mathematical Model." Biotechnology and Bioengineering **55**(4): 592-608.

Savageau, M. A. (1969). "Biochemical systems analysis. I. Some mathematical propeties of the rate law for the component enzymatic reactions." J. Theor. Biol. **25**: 365-369.

Schmid, J. W., K. Mauch, M. Reuss, E. D. Gilles and A. Kremling (2004). "Metabolic design based on a coupled gene expression-metabolic network model of tryptophan production in Escherichia coli." Metab Eng **6**(4): 364-77.

Schuster, R. and H. G. Holzhutter (1995). "Use of mathematical models for predicting the metabolic effect of large-scale enzyme activity alterations. Application to enzyme deficiencies of red blood cells." Eur J Biochem **229**(2): 403-18.

Shuler, M. L. (1999). "Single-cell models: promise and limitations." J Biotechnol **71**(1-3): 225-8.

Shuler, M. L., S. Leung and C. C. Dick (1979). "A mathematical model for the growth of a single bacterial cell." Ann. NY Acad. Sci. **326**: 35-55.

Small, R. and H. Kacser (1994). "A method for increasing the concentration of a specific internal metabolite in steady-state systems." Eur J Biochem **226**: 649-656.

Stelling, J., S. Klamt, K. Bettenbrock, S. Schuster and E. D. Gilles (2002). "Metabolic network structure determines key aspects of functionality and regulation." Nature **420**(6912): 190-3.

Stephanopoulos, G. and T. W. Simpson (1997). "Flux amplification in comlex metabolic networks." Chemical Engineering Science **52**(15): 2607-2627.

Stephanopoulos, G. and J. J. Vallino (1991). "Network rigidity and metabolic engineering in metabolite overproduction." Science **252**(5013): 1675-81.

Stephanopoulos, G. N., A. A. Aristidou and J. Nielsen (1998). Metabolic Engineering. Principles and Methods. New York, Academic Press.

ter Kuile, B. H. and H. V. Westerhoff (2001). "Transcriptome meets metabolome: hierarchical and metabolic regulation of the glycolytic pathway." FEBS Lett **500**(3): 169-71.

Tomita, M. (2001). "Whole-cell simulation: a grand challenge of the 21st century." Trends Biotechnol. **19**(6): 205-210.

Underwood, S. A., M. L. Buszko, K. T. Shanmugam and L. O. Ingram (2002). "Flux through citrate synthase limits the growth of ethanologenic Escherichia coli KO11 during xylose fermentation." Appl Environ Microbiol **68**(3): 1071-81.

Underwood, S. A., M. L. Buszko, K. T. Shanmugam and L. O. Ingram (2004). "Lack of protective osmolytes limits final cell density and volumetric productivity of ethanologenic Escherichia coli KO11 during xylose fermentation." Appl Environ Microbiol **70**(5): 2734-40.

Visser, D., J. W. Schmid, K. Mauch, M. Reuss and J. J. Heijnen (2004). "Optimal re-design of primary metabolism in *Escherichia coli* using linglog kinetics." Biotechnology and Bioengineering.

Walsh, K. and D. E. Koshland, Jr. (1985). "Characterization of rate-controlling steps in vivo by use of an adjustable expression vector." Proc Natl Acad Sci U S A **82**(11): 3577-81.

Weitzman, P. D. (1966). "Regulation of citrate synthase activity in escherichia coli." Biochim Biophys Acta **128**(1): 213-5.

TABLES

Table 1. Double index (*j*:*s*) relates generalized regulatory parameter α_s to regulated reaction rate r_j.

№	Enzyme	Regulation	Notation
1	PTS	inhibition by g6p	(1:31)
2	PGI	inhibition by 6pg	(2:32)
3	PFK	inhibition by pep	(3:33)
4	PFK	activation by adp	(3:34)
5	PFK	activation by amp	(3:35)
6	PK	activation by amp	(10:36)
7	PK	activation by fdp	(10:37)
8	PK	inhibition by atp	(10:38)
9	G1PAT	activation by nadph	(14:39)
10	G6PDH	inhibition by nadph	(21:40)
11	PGDH	inhibition by atp	(22:41)
12	PGDH	inhibition by nadph	(22:42)
13	PEPCxylase	activation by fdp	(12:43)

Table 2. Best ratios of steady state fluxes in the optimized and original models correspond to (*i*) regulation alterations, (*ii*) enzyme level modulations, and (*iii*) both regulation and enzyme level manipulations.

Flux	**Regulation**	**Enzyme Level**	**Enzyme Level & Regulation**
PTS	1.43	3.16	8.15
Serine	1.057	20.913	22.615

Table 3. Control coefficients for regulated enzymes at the original steady state, EEC –Enzyme Elasticity Coefficient, FCC – Flux Control Coefficient, and FRC – Flux Response Coefficient. Notation *j*:*s* relates regulatory parameter α_s to reaction *j* (i.e., rate r_j depending on parameter α_s).

№	Enzyme	Modifier	Regulation	EEC	**FCC** (PTS)	**FRC** (PTS)	**FCC** (Serine)	**FRC** (Serine)
1:31	PTS	g6p	negative	-0.978	0.416	-0.407	0.192	-0.187
2:32	PGI	6pg	negative	-0.551	0.000692	-0.000381	0.000374	-0.000206
3:33	PFK	pep	negative	-2.047	0.242	-0.496	0.14	-0.287
3:34	PFK	adp	*positive*	0.025	0.242	0.00605	0.14	0.0035
3:35	PFK	amp	*positive*	0.268	0.242	0.065	0.14	0.0376
10:36	PK	amp	*positive*	0.000226	0.0109	0.00000246	-0.122	-0.0000275
10:37	PK	fdp	*positive*	0.0000682	0.0109	0.00000074	-0.122	-0.00000831
10:38	PK	atp	negative	-0.0000544	0.0109	-0.0000006	-0.122	0.00000664
14:39	G1PAT	fdp	*positive*	0.731	0.00721	0.00527	-0.00934	-0.00683
21:40	G6PDH	nadph	negative	-0.419	0.115	-0.0483	-0.0721	0.0302
22:41	PGDH	atp	negative	-0.012	0.000389	-0.00000464	0.000211	-0.0000025
22:42	PGDH	nadph	negative	-0.485	0.000389	-0.000189	0.000211	-0.000102
12:34	PEPxylase	fdp	*positive*	0.019	0.0387	0.00073	-0.126	-0.00238

Table 4. Alternative best optimal enzyme/regulation subsets leading to the increased PTS rate. Indices highlighted in bold correspond to enzymes exerting high control on the PTS rate. Signature (+)/(-) corresponds to the increase/decrease in the corresponding enzyme property (i.e., the enzyme level or enzyme regulation), respectively.

Size	Enzyme Set	Flux Ratio	Enzyme/Regulation Set	Flux Ratio
1	**1(+)**	1.073	**1:31(-)**	1.08
2	**1(+) 15(+)**	1.233	**1:31(-)** 14:39(+)	1.462
3	**1(+) 6(+) 12(+)**	1.628	**1:31(-) 6(+)** 14:39(+)	2.173
4	**1(+) 3(+) 6(+) 12(+)**	2.246	**1:31(-) 3:33(-) 6(+) 12:43(+)**	3.88
5	**1(+) 3(+) 6(+) 11(+) 12(+)**	2.541	**1:31(-) 3:33(-)** 5(+) **6(+) 12:43(+)**	4.42
6	**1(+) 3(+)** 5(+) **6(+) 11(+) 12(+)**	2.843	**1:31(-) 3:33(-)** 5(+) **6(+) 11(+) 12:43(+)**	5.55
7	**1(+) 3(+)** 5(+) **6(+)** 7(-) **11(+) 12(+)**	2.892	**1:31(-) 3:33(-)** 4(+) 5(+) **6(+) 11(+) 12:43(+)**	6.88
8	**1(+) 3(+)** 4(-) 5(+) **6(+)** 7(-) **11(+) 12(+)**	2.964	**1:31(-) 3:33(-)** 4(+) 5(+) **6(+)** 7(+) **11(+) 12:43(+)**	7.1
9	**1(+) 3(+)** 4(-) 5(+) **6(+)** 7(-) 8(-) **11(+) 12(+)**	3.048	**1:31(-) 3:33(-)** 4(+) 5(+) **6(+)** 7(+) 8(-) **11(+) 12:43(+)**	7.1
10	**1(+) 3(+)** 4(-) 5(+) **6(+)** 7(-) 8(-) 9(-) **11(+) 12(+)**	3.155	**1:31(-) 3:33(-)** 4(+) 5(+) **6(+)** 7(+) 8(-) 9(-) **11(+) 12:43(+)**	7.44

Table 5. Alternative best optimal enzyme/regulation subsets leading to the increased serine production flux. Indices highlighted in bold correspond to enzymes exerting high control on the serine flux. Signature (+)/(-) corresponds to the increase/decrease in the corresponding enzyme property (i.e., the enzyme level or enzyme regulation), respectively.

Size	Enzyme Set	Flux Ratio	Enzyme/Regulation Set	Flux Ratio
1	**17(+)**	1.88	**21:40(+)**	1.034
2	**1(+) 17(+)**	4.652	**1:31(-) 17(+)**	4.754
3	**1(+) 6(+) 17(+)**	9.086	**1:31(-) 6(+) 17(+)**	9.963
4	**1(+) 3(+) 6(+) 17(+)**	14.451	**1:31(-) 3:35(+) 6(+) 17(+)**	16.65
5	**1(+) 3(+) 6(+) 11(+) 17(+)**	15.933	**1:31(-) 3:35(+)** 5(+) **6(+) 17(+)**	18.518
6	**1(+) 3(+)** 5(-) **6(+) 11(+) 17(+)**	17.418	**1:31(-) 3:35(+)** 5(+) **6(+) 11(+) 17(+)**	20.039
7	**1(+) 3(+)** 4(-) 5(-) **6(+) 11(+) 17(+)**	19.085	**1:31(-) 3:35(+)** 4(-) 5(+) **6(+) 11(+) 17(+)**	22.012
8	**1(+) 3(+)** 4(-) 5(-) **6(+)** 7(-) **11(+) 17(+)**	19.838	**1:31(-) 3:35(+)** 4(-) 5(-) **6(+)** 7(-) **11(-) 17(+)**	22.371
9	**1(+)** 2(-) **3(+)** 4(-) 5(-) **6(+)** 7(-) **11(+) 17(+)**	20.538	**1:31(-)** 2:32(-) **3:35(+)** 4(-) 5(-) **6(+)** 7(-) **11(-) 17(+)**	22.615
10	**1(+)** 2(-) **3(+)** 4(-) 5(-) **6(+)** 7(-) 9(-) **11(+) 17(+)**	20.591		

FIGURE CAPTIONS

Figure 1. *Escherichia coli* central carbon metabolism.

Figure 2. A schematic simulated annealing pseudo-code representation.

Figure 3. Shown are the best optimal reaction rate ratios (i.e., r/r_0) for (*a*) PTS and (*b*) SerSynth, plotted as a function of size *D* of modulated enzyme/regulation subsystems (i.e., *D* = 1, 2, …, 10). Here *r* and r_0 are the reaction rates in the optimized and original models, respectively. White circles correspond to the case where enzyme levels are modulated while enzyme regulation is kept unchanged. Solid circle correspond to the case where both enzyme levels and enzyme regulations are manipulated. In both cases, the corresponding optimal steady state solutions are *stable*. Gray diamonds in case (*a*) correspond to the unstable optimal stationary solutions near a Hopf bifurcation. No stable optimal solution in case (*b*) for D = 10 could be found.

Figure 4. Flux Control Coefficients (FCCs) for the PTS reaction (white bars) and serine production (solid bars), respectively.

Figure 5. (*a*) Flux Control Coefficients (FCCs) for the PTS reaction. (*b*) Relative enzyme levels at the optimal stable steady states. White bars correspond to the original non-perturbed case as shown in Figure 4. Gray bars correspond to the case where the enzyme levels are optimally chosen in the absence of any regulation alteration. Black bars correspond to the case where regulatory structures for enzymes PTS and PFK are altered and levels of enzymes TIS, GAPDH, PDH, and PEPxylase are modulated.

Figure 6. The distributions of stable steady state fluxes relative to the PTS flux at the origin reference non-perturbed steady state. Values of fluxes highlighted in bold correspond to the case where regulatory structures for enzymes PTS and PFK, and levels of enzymes TIS, GAPDH, PDH, and PEPxylase are simultaneously manipulated. The other (i.e., non-highlighted) values correspond to the case where the levels of all the six enzymes are modulated in the absence of any regulatory changes.

Figure 7. (*a*) Flux Control Coefficients (FCCs) for the PTS reaction. (*b*) Relative enzyme levels at the optimal stable steady states. White bars correspond to the original non-perturbed case as shown in Figure 4. Gray bars correspond to the case where the enzyme levels are optimally chosen in the absence of any regulation alteration. Black bars correspond to the case where regulatory structures for enzymes PTS and PFK are altered and levels of enzymes TIS, GAPDH, PDH, and SerSynth are simultaneously modified.

Figure 8. The distributions of stable steady state fluxes relative to the PTS flux at the origin reference state. Values of fluxes highlighted in bold correspond to the case where regulatory structures for enzymes PTS and PFK, and levels of enzymes TIS, GAPDH, PDH, and SerSynth are simultaneously modified. The other values correspond to the case where the levels of all the six enzymes are modulated in the absence of any regulatory changes.

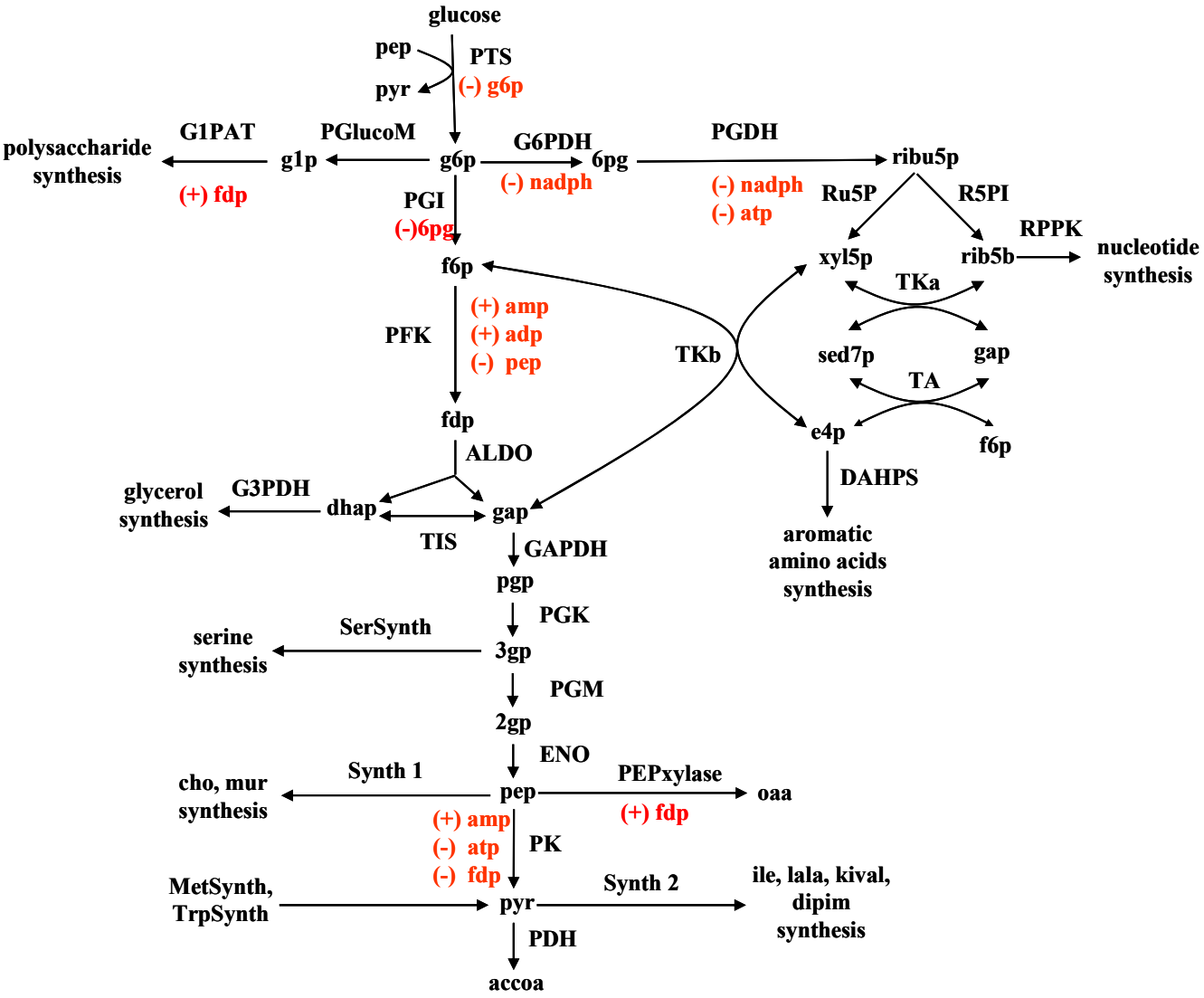
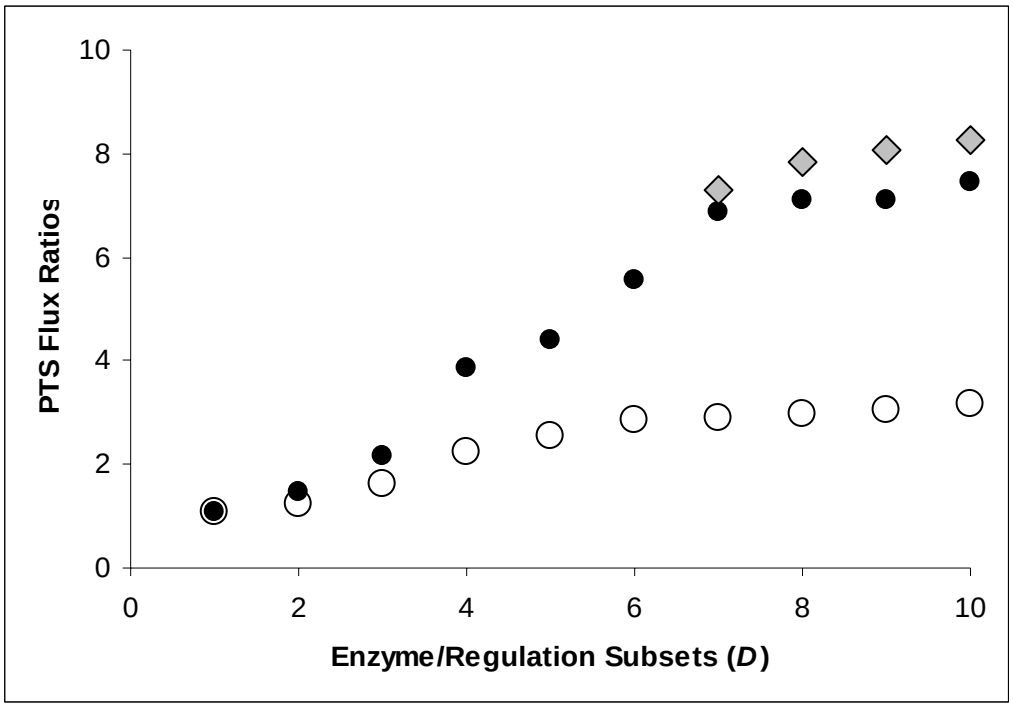

glucose
pep
pyr
PTS
(-) g6p
polysaccharide synthesis
G1PAT
(+) fdp
PGlucoM
g1p
g6p
G6PDH
(-) nadph
6pg
PGDH
(-) nadph
(-) atp
ribu5p
Ru5P
R5PI
PGI
(-)6pg
f6p
xyl5p
rib5b
RPPK
nucleotide synthesis
TKa
PFK
(+) amp
(+) adp
(-) pep
TKb
sed7p
gap
TA
fdp
e4p
f6p
ALDO
DAHPS
glycerol synthesis
G3PDH
dhap
gap
TIS
GAPDH
aromatic amino acids synthesis
pgp
PGK
serine synthesis
SerSynth
3gp
PGM
2gp
ENO
cho, mur synthesis
Synth 1
pep
PEPxylase
oaa
(+) fdp
(+) amp
(-) atp
(-) fdp
PK
MetSynth, TrpSynth
pyr
Synth 2
ile, lala, kival, dipim synthesis
PDH
accoa

Figure 1

1. Generate an initial enzyme/regulation subsystem S_0
2. Set $S_b = S_c = S_t = S_0$
3. $r_b = r_c = r_t =$ Optimize(S_t)
4. `for` $i = 1{:}MaxIter$
5. $S_t =$ Select or Terminate (S_c)
6. $r_t =$ Optimize (S_t)
7. `if` $r_t > r_b$
8. $S_b = S_c = S_t$
9. $r_b = r_c = r_t$
10. `else`
11. $anneal = e^{(r_t - r_c)/T}$
12. Generate a random $d \in (0,1)$
13. `if` $d < anneal$
14. $S_c = S_t$
15. $r_c = r_t$
16. `end if`
17. `end if`
18. `if` $[i/J]$
19. $= 0$
20. $T = \beta \cdot T$
21. `end if`
22. `end loop`

Figure 2

(a)


PTS Flux Ratios
Enzyme/Regulation Subsets (D)
0
2
4
6
8
10

(*b*)

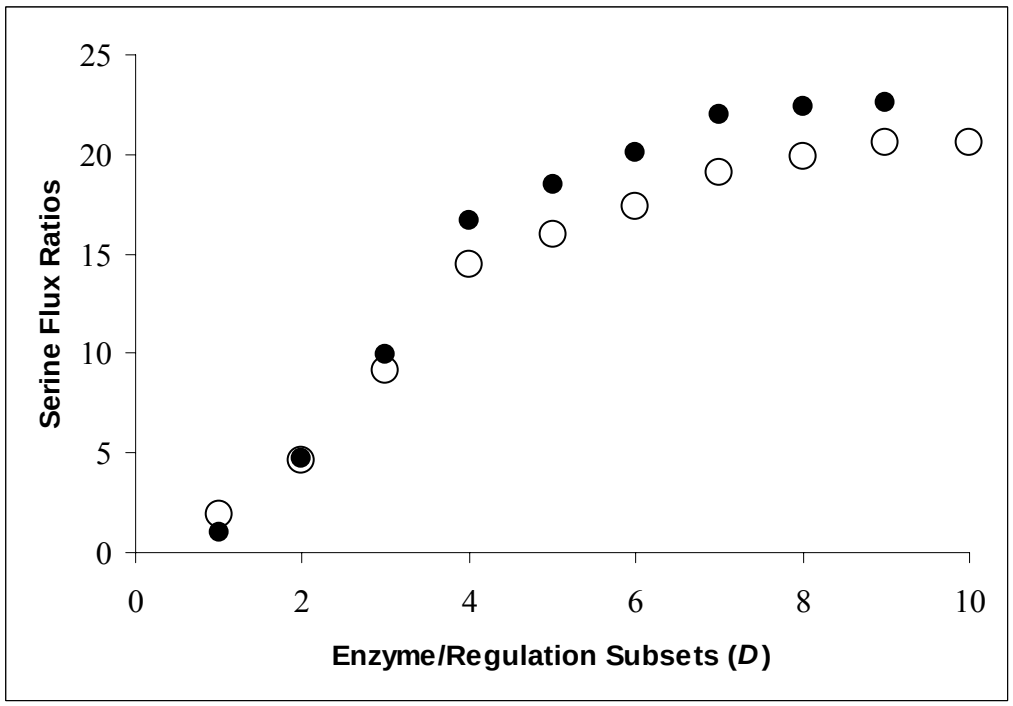

Serine Flux Ratios
Enzyme/Regulation Subsets (D)
0
5
10
15
20
25
0
2
4
6
8
10

Figure 3

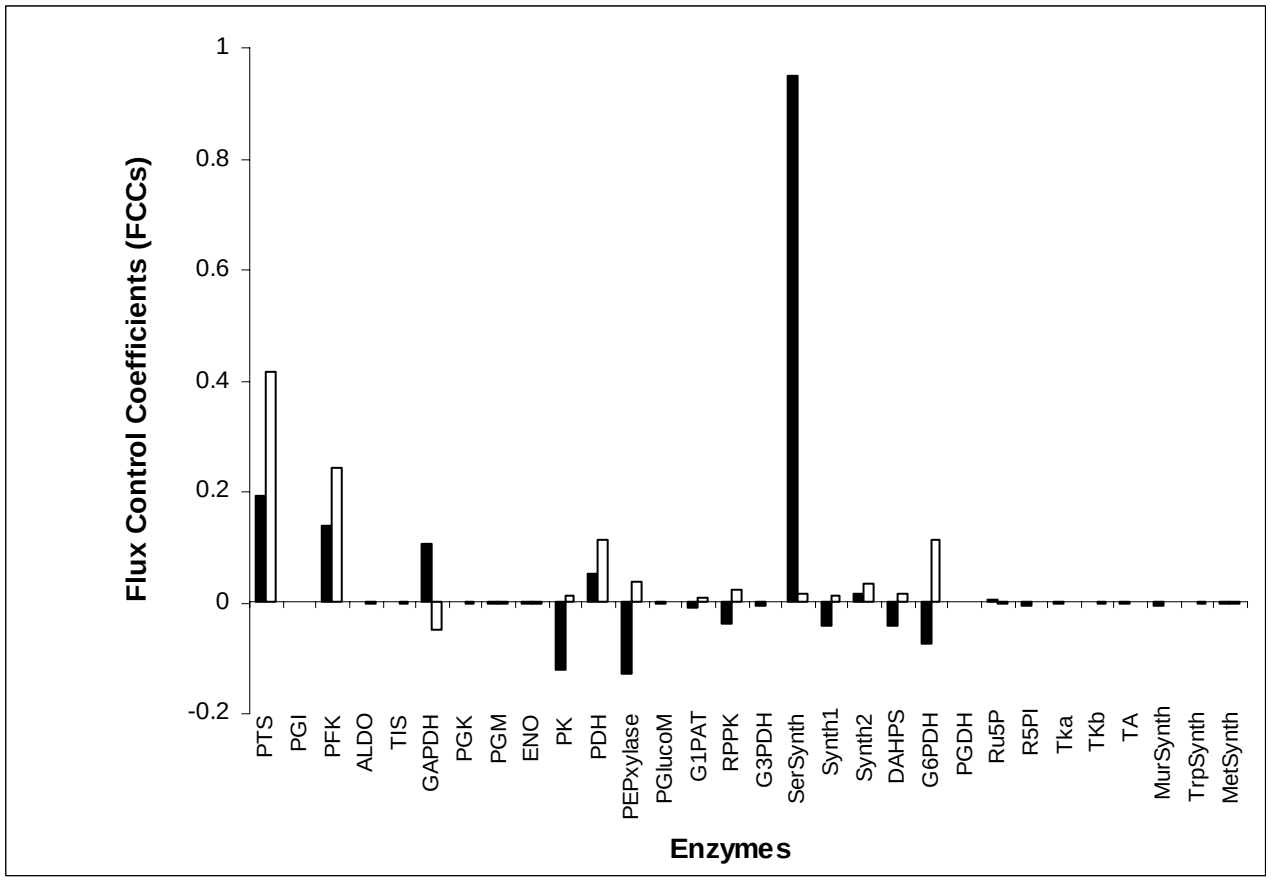
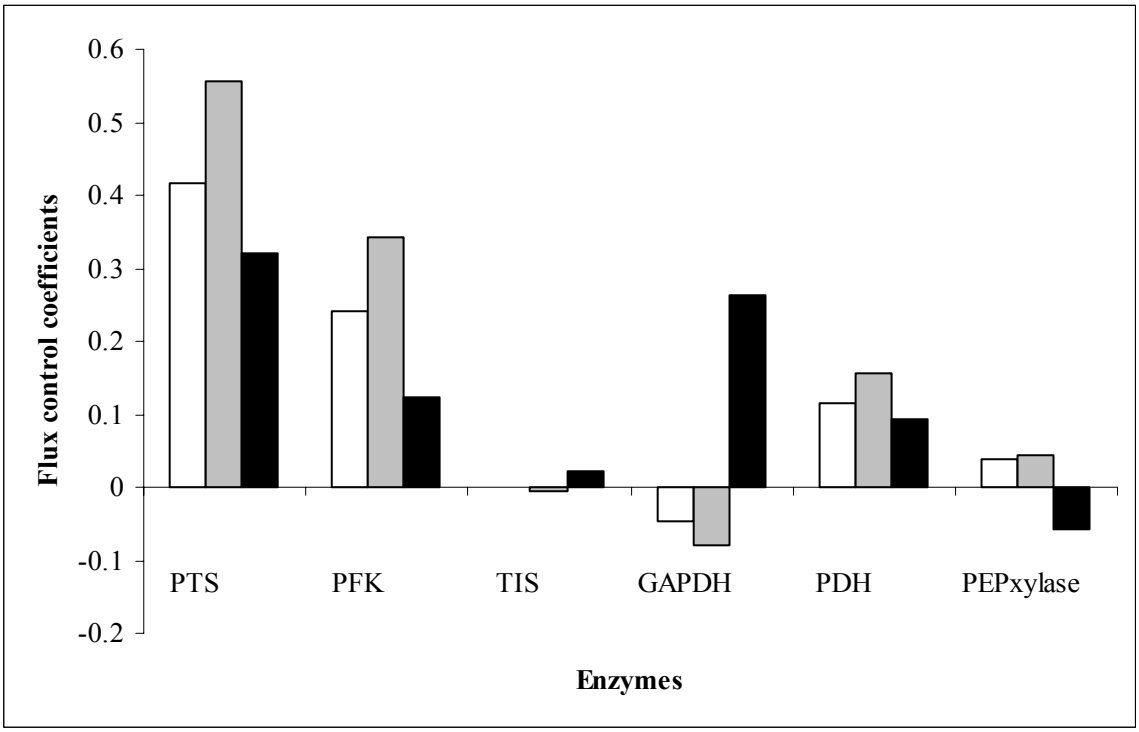

Flux Control Coefficients (FCCs)
1
0.8
0.6
0.4
0.2
0
-0.2
PTS
PGI
PFK
ALDO
TIS
GAPDH
PGK
PGM
ENO
PK
PDH
PEPxylase
PGlucoM
G1PAT
RPPK
G3PDH
SerSynth
Synth1
Synth2
DAHPS
G6PDH
PGDH
Ru5P
R5PI
Tka
TKb
TA
MurSynth
TrpSynth
MetSynth
Enzymes

Figure 4

(*a*)


0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
Flux control coefficients
PTS
PFK
TIS
GAPDH
PDH
PEPxylase
Enzymes

(*b*)

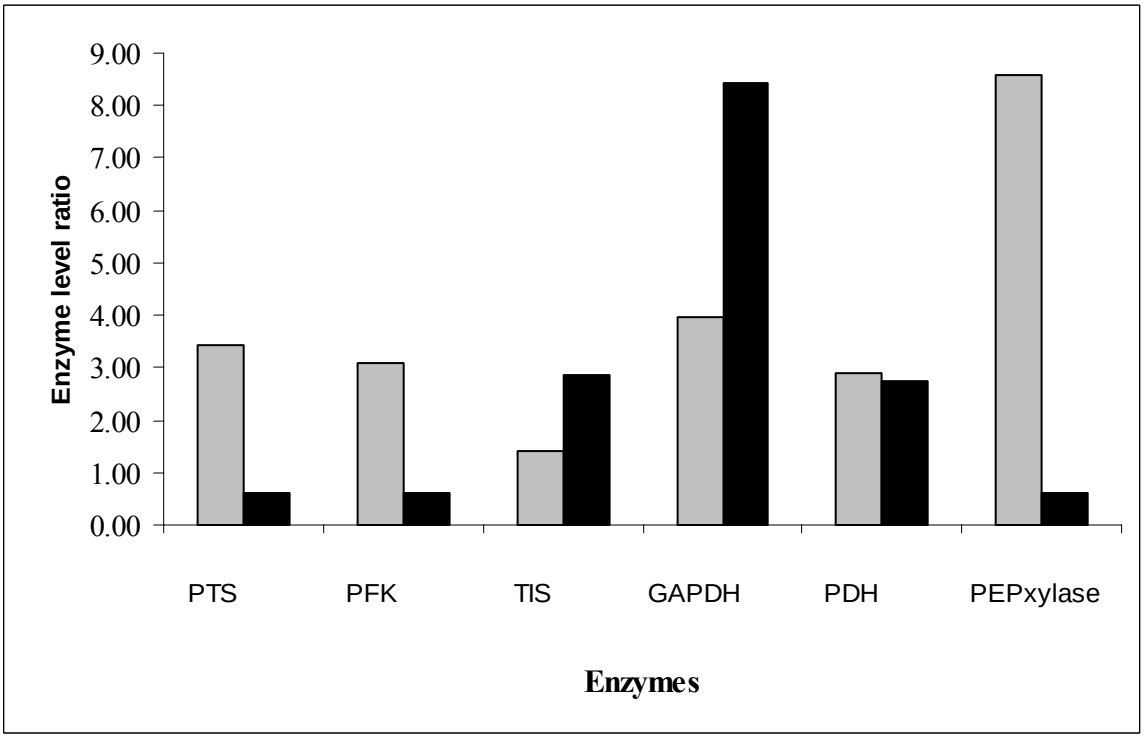

9.00
8.00
7.00
6.00
5.00
4.00
3.00
2.00
1.00
0.00
Enzyme level ratio
PTS
PFK
TIS
GAPDH
PDH
PEPxylase
Enzymes

Figure 5

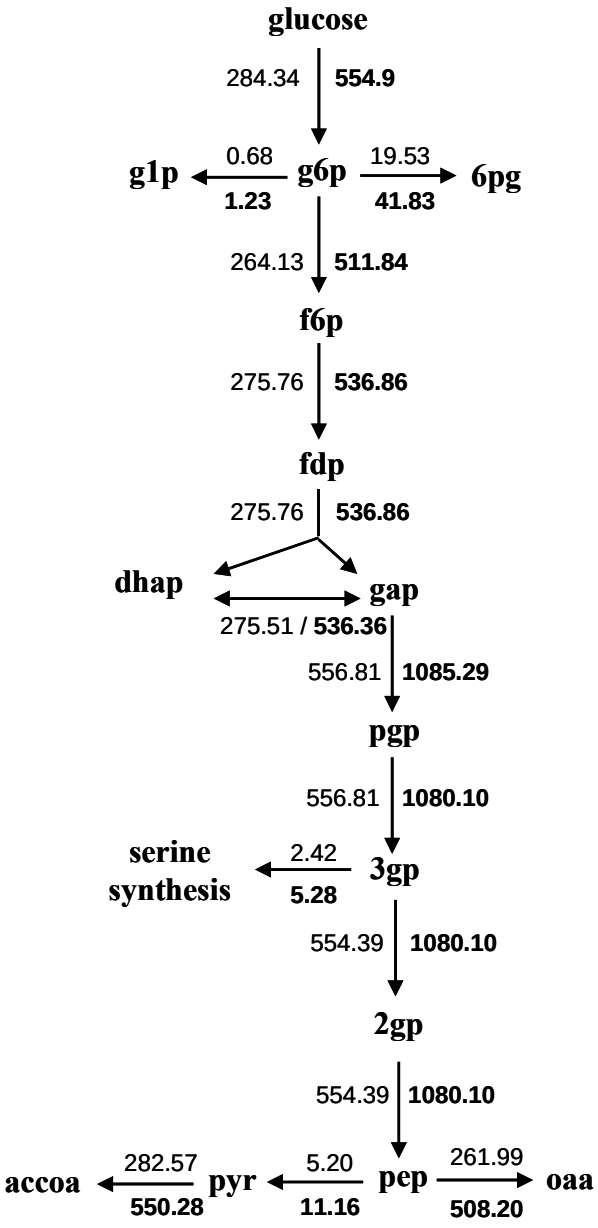
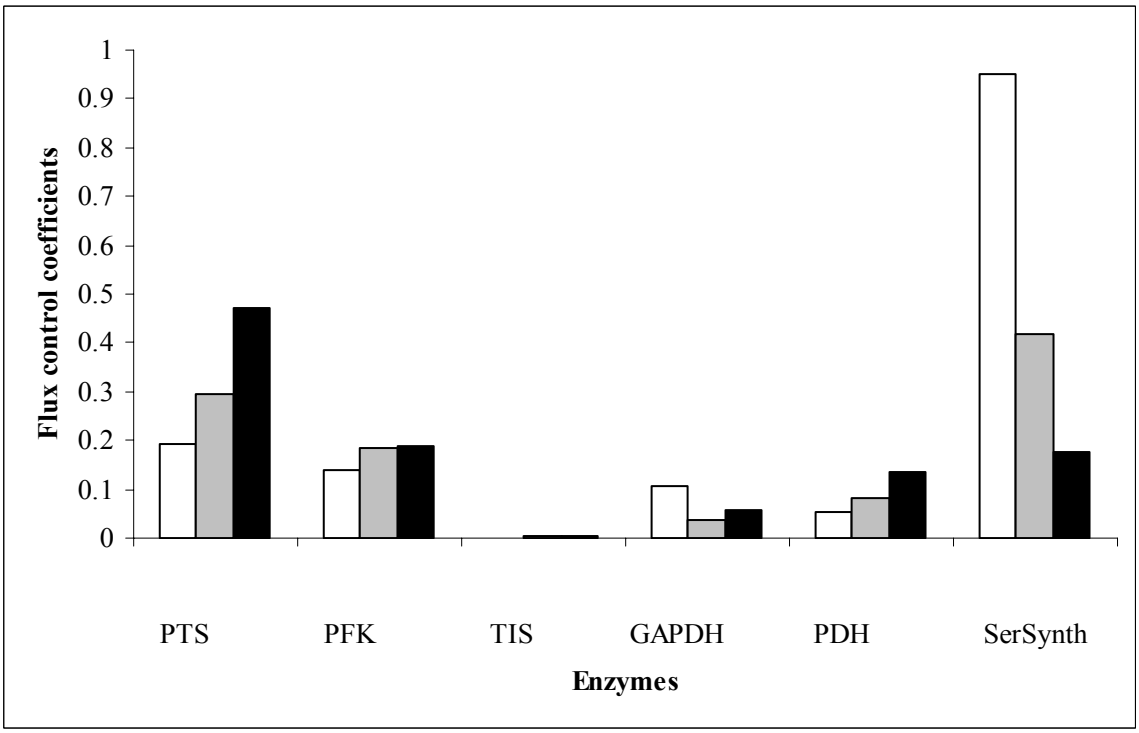

glucose
284.34
554.9
g1p
0.68
1.23
g6p
19.53
41.83
6pg
264.13
511.84
f6p
275.76
536.86
fdp
275.76
536.86
dhap
gap
275.51 / 536.36
556.81
1085.29
pgp
556.81
1080.10
serine synthesis
2.42
5.28
3gp
554.39
1080.10
2gp
554.39
1080.10
accoa
282.57
550.28
pyr
5.20
11.16
pep
261.99
508.20
oaa

Figure 6

(a)


Flux control coefficients
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
PTS
PFK
TIS
GAPDH
PDH
SerSynth
Enzymes

(b)

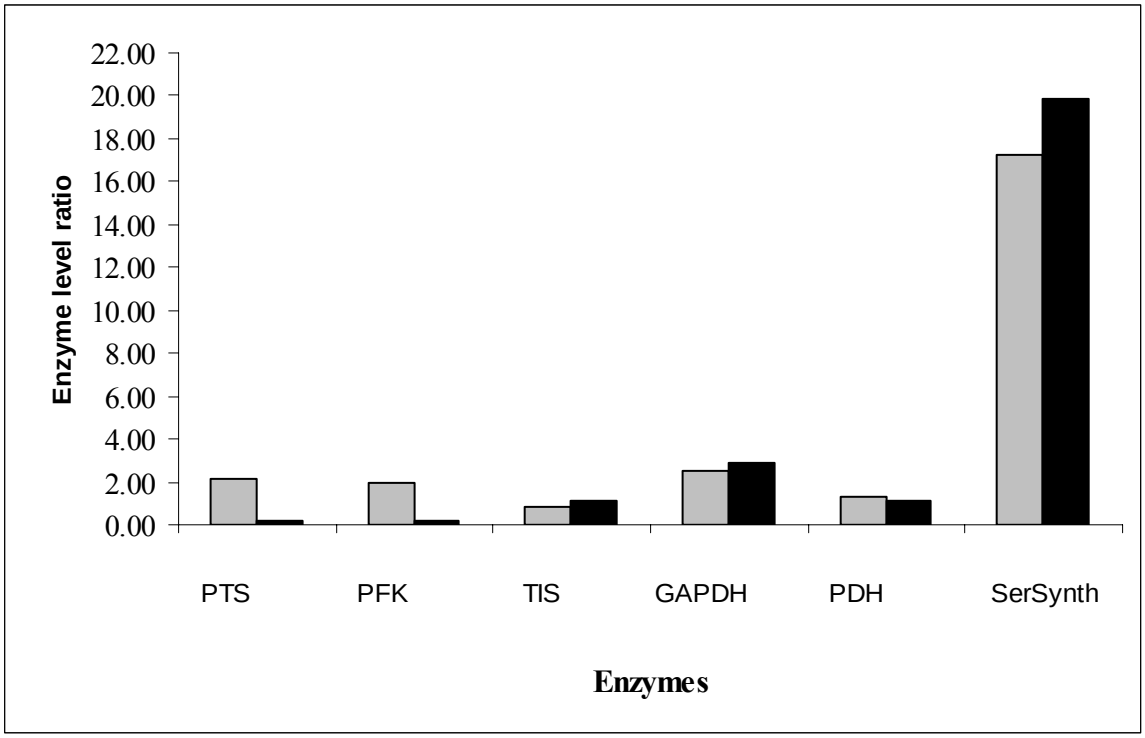

Enzyme level ratio
22.00
20.00
18.00
16.00
14.00
12.00
10.00
8.00
6.00
4.00
2.00
0.00
PTS
PFK
TIS
GAPDH
PDH
SerSynth
Enzymes

Figure 7

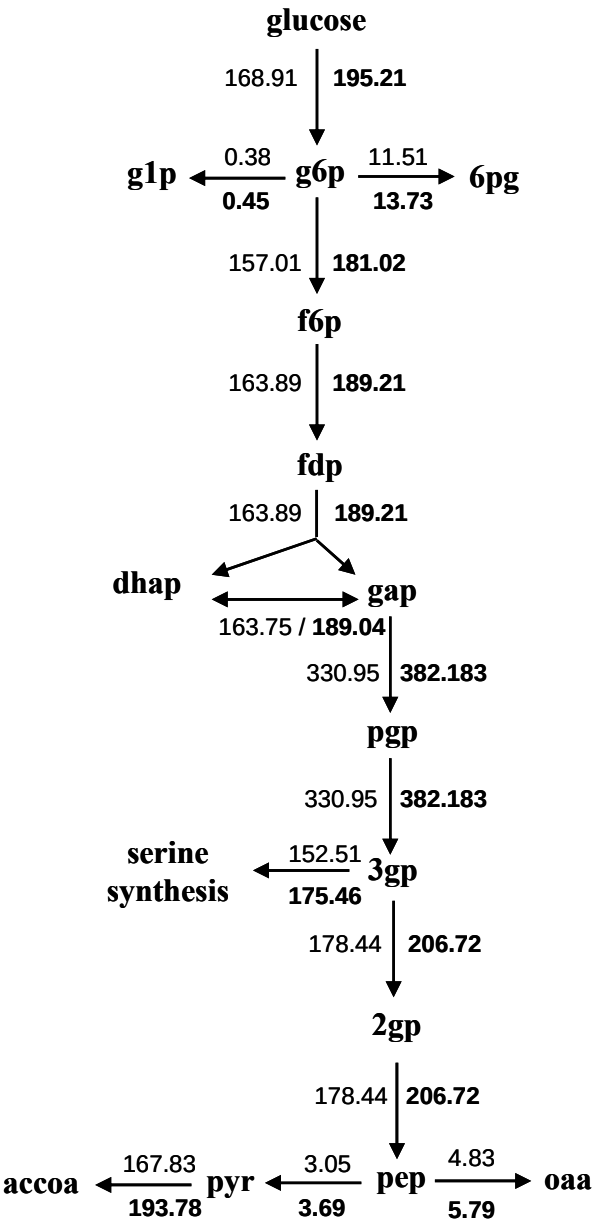

glucose
168.91
195.21
g1p
0.38
0.45
g6p
11.51
13.73
6pg
157.01
181.02
f6p
163.89
189.21
fdp
163.89
189.21
dhap
gap
163.75 / 189.04
330.95
382.183
pgp
330.95
382.183
serine synthesis
152.51
175.46
3gp
178.44
206.72
2gp
178.44
206.72
accoa
167.83
193.78
pyr
3.05
3.69
pep
4.83
5.79
oaa

Figure 8